The National Park Service

Inventory & Monitoring

Northern Colorado Plateau
Vital Signs Network and Prototype Cluster

Data Management Plan

September, 2005

Prepared by:

Margaret Beer, Data Manager
Elizabeth Nance, Biologist
Aneth Wight, GIS Technician
Melissa Powell, Biological Technician
Russ DenBleyker, IT Specialist

❖

National Park Service
Inventory and Monitoring Program
Northern Colorado Plateau Network
2282 S. West Resource Blvd.
Moab, Utah 84532-3298

Suggested citation:

Beer, M., E. Nance, A. Wight, M. Powell, and R. DenBleyker. 2005. Northern Colorado Plateau Network, data management plan. National Park Service, Inventory and Monitoring Program, Northern Colorado Plateau Network, Moab, UT. 86 p. plus appendices.

TABLE OF CONTENTS

Figures

Tables

Appendices

Acknowledgements

Many people have contributed to the production of this plan. The overall plan outline and components of all chapters were developed by the Inventory and Monitoring Program Data Management Planning Workgroup, ably headed by John Boetsch and comprising data managers from networks across the country. Special thanks are in order to chapter coordinators Doug Wilder, Rob Daley, John Boetsch, Patrick Flaherty, Geoff Sanders, Debbie Angell, Teresa Leibfried, Gareth Rowell, and Sara Stevens. Vel Decker provided invaluable review and suggestions, and Dorothy Mortenson and Nicole Tancreto have been an ongoing source of ideas and assistance. All workgroup members contributed their thoughts and materials generously and without concern for credit. The collaboration of network data managers has been a highlight of producing this plan, and sets a precedent for cooperation in the data management tasks we all face in the upcoming years.

We thank the national-level Inventory and Monitoring Program data management team, who have provided the vision and created the tools that are essential to our work. In particular, Lisa Nelson, Danelle Malget, Chris Dietrich, Willene Hendon, Joe Gregson, Wendy Schumacher, Mark Wotawa, Simon Kingston, and Alison Loar have all provided outstanding technical support, and have been patient and open-minded to our questions and suggestions.

We are fortunate to have Vicki Webster, curator at the Southeast Utah Group, in the same offices as the Northern Colorado Plateau Network. Vicki has educated us all on the value of setting high curatorial standards, and has helped our program better understand the significant intersection of NPS museum management and the Inventory and Monitoring Program.

Finally, we thank Steve Fancy for his unwavering commitment to data management in the National Park Service.

Executive Summary

Information is the common currency among the activities and staff involved in natural resource management in the National Park Service (NPS). The central mission of the National Park Service's Inventory and Monitoring (I&M) Program is to acquire, manage, analyze, and distribute scientific information on the status and trends of specific park natural resources. Intended users of this information include park managers, cooperators, researchers, and the general public.

A cornerstone of the Inventory and Monitoring Program is the strong emphasis placed on data management. All I&M networks, including the Northern Colorado Plateau Network (NCPN), expect to invest at least thirty percent of their available resources in data management, analysis, and reporting activities.

The goal of the Northern Colorado Plateau Network's data management program is to maintain, in perpetuity, the ecological data and related analyses that result from the network's resource inventory and monitoring work. The purpose of the Data Management Plan is to describe the resources and processes required to ensure the following standards for data acquired or managed by NCPN:

- *Accuracy*: The quality of the data collected and managed by the I&M Program is paramount. Analyses performed to detect ecological trends or patterns require data with minimal error and bias. Inconsistent or poor-quality data can limit the detectibility of subtle changes in ecosystem patterns and processes, lead to incorrect interpretations and conclusions, and could greatly compromise the credibility and success of the I&M Program. To ensure that NCPN produces and maintains data of the highest possible quality, procedures are established to identify and minimize errors at each stage of the data lifecycle.

- *Security*: Digital and hard-copy data must be maintained in environments that protect against loss, either due to electronic failure or to poor storage conditions. NCPN digital data are stored in multiple formats on a secure server, and are part of an integrated backup routine that includes rotation to off-site storage locations. In addition, NCPN is working with NPS museum curators and archivists to ensure that related project materials such as field notes, data forms, specimens, photographs, and reports are properly cataloged, stored, and managed in archival conditions.

- *Longevity*: Countless data sets have become unusable over time either because the format is outdated (e.g., punchcards), or because metadata is insufficient to determine the data's collection methods, scope and intent, quality assurance procedures, or format. While proper storage conditions, backups, and migration of data sets to current platforms and software standards are basic components of data longevity, comprehensive data documentation is equally important. NCPN uses a suite of metadata tools to ensure that data sets are consistently documented, and in formats that conform to current federal standards.

- *Usability*: One of the most important responsibilities of the Inventory and Monitoring Program is to ensure that data collected, developed, or assembled by

NCPN staff and cooperators are made available for decision-making, research, and education. Providing well-documented data in a timely manner to park managers is especially important to the success of the program. NCPN must ensure that:

- data can be easily found and obtained
- data are subjected to full quality control before release
- data are accompanied by complete metadata
- sensitive data are identified and protected from unauthorized access and distribution

NCPN's main mechanism for distribution of the network's inventory and monitoring data will be the Internet, which will allow data and information to reach a broad community of users. As part of the NPS I&M Program, web-based applications and repositories have been developed to store a variety of park natural resource information (Table 1).

Table 1. Data that are provided on the NCPN and national I&M websites

Web Application Name	Data available at site
NPSpecies	Database of vascular plant and vertebrate species known or suspected to occur on NPS park units (NPSpecies Home Page).
NatureBib	Bibliography of park-related natural resource information (NatureBib Home Page
NPSFocus	Portal to a variety of NPS information sources; will include NatureBib and NR/GIS Data Store links
Biodiversity Data Store	Digital archive of documents, GIS datasets and non-GIS dataset files that document the presence/absence, distribution and/or abundance of taxa in National Park Service units (Biodiversity Service Center Home Page)
NR-GIS Data Store	Park-related metadata and selected data sets (spatial and non-spatial) — (NR-GIS Metadata and Data Store Home Page)
NCPN Website	Reports and metadata for NCPN projects; certified species lists; search and reporting tools for data; data downloads; database templates (NCPN Home Page)

The Northern Colorado Plateau Network's information acquires its real value when it reaches those who can apply it. If these web portals do not meet a specific user's requirements, NCPN data management staff will work with users on an individual basis to ensure receipt of the desired information in the requested format.

Data Management Roles and Responsibilities

Data management is collaborative work that involves many persons with a broad range of expertise and abilities. All network staff have a role in data stewardship, and project data sets and products reflect all who have contributed.

Table 2 lists data-related roles and primary responsibilities, from field-based data collection, to final distribution and archiving. The fundamental role of the network data manager is to coordinate these tasks.

Table 2. Roles and primary responsibilities related to network data management.

Role	Primary responsibilities related to data management
Project crew member	Collect, record, verify data; perform data entry; organize field forms, photos, other related materials
Project crew leader	Supervise crew; communicate regularly with data manager and project leader
GIS specialist	Oversee GPS data collection; manage spatial data; prepare maps; perform spatial analyses
IT specialist	Apply database and programming skills to network projects; maintain information systems to support data management
Project leader	Direct operations, including data management requirements, for network projects
Resource specialist	Evaluate validity and utility of project data; document, analyze, publish data and associated information products
Network data manager	Ensure program data are organized, useful, compliant, safe, and available
Quantitative ecologist	Determine project objectives and sample design; perform and document data analysis and synthesis; prepare reports
Network coordinator	Coordinate and oversee all network activities
Park or regional curator	Ensure project results (documents, specimens, photographs, etc.) are cataloged andaccessioned into NPS or other repositories
I&M data manager (national level)	Provide service-wide database support and services; provide data management coordination among networks
End users (managers, scientists, interpreters, public)	Inform and direct the scope of science information needs; interpret information and use to direct or support decisions

Data Sources and Priorities

There are multiple sources of significant data related to natural resources in NCPN parks. The types of work that may generate these data include:

- inventories
- monitoring
- protocol development pilot studies
- special-focus studies performed by internal staff, contractors or cooperators
- external research projects
- studies performed by other agencies on park or adjacent lands
- resource impact evaluations related to park planning and compliance
- resource management and restoration work.

Prioritizing data management efforts in a sea of unmanaged data

- Highest priority is to produce and curate high-quality, well-documented data originating with the Inventory and Monitoring Program
- As time and resources permit, assist with data management for current projects, legacy data, and data originating outside the Inventory and Monitoring Program that complement program objectives
- In addition, help ensure good data management practices for park-based natural resource projects that are just beginning to be developed and implemented.

Because the I&M Program focuses on natural resource inventories and long-term monitoring, NCPN's first data management priority is the data and information that results from these efforts. However, the standards, procedures, and approaches to data management developed by NCPN are being applied to other natural resource data sources.

For example, all natural resource parks need a basic suite of resource inventory data in order to manage their resources and support a successful monitoring program. The national Inventory and Monitoring Program has determined that a minimum of 12 inventory data sets, including both biotic and abiotic components, will be acquired by all parks. NCPN is working with individual parks and national NPS programs to acquire and standardize these basic resource data sets, and make them widely available. The data sets are:

- Natural resource bibliography
- Base cartographic data
- Geology map
- Soils map
- Weather data
- Air quality data
- Location of air quality monitoring stations
- Water body location and classification
- Water quality data
- Vegetation map
- Species distribution and status of vertebrates and vascular plants
- Documented species list of vertebrates and vascular plants

A summary of the status of these data sets for network parks is presented in Appendix J (Status of the 12 natural resource inventories, Northern Colorado Plateau Network).

Data Management Categories

Data from park and network sources can generally be placed in the following data management categories:

1. *Data managed in service-wide databases.*
NCPN uses three databases developed by the I&M Washington Office (WASO). NatureBib is a bibliographic tool for cataloging reports, publications, or other documents that relate to natural resources in park units. Dataset Catalog is used to document primarily non-spatial natural resource-related databases or other data assemblages (e.g., photographs, field data sheets). NPSpecies is used by the network to develop and maintain lists of vertebrates and vascular plants in network parks, along with associated supporting evidence.

2. *Data developed or acquired directly by the network as a result of inventory, monitoring, or other projects, and managed by NCPN.*
This category includes project-related protocols, field data, reports, spatial data, and associated materials such as field forms and photographs provided to NCPN by contractors or developed by NCPN staff. Projects can be short-term (one to three years duration) or long-term (ongoing monitoring).

3. *Data that, while not developed or maintained by NCPN, are used as primary data sources or provide context to other data sets.*
Examples of this category include: geographic information system (GIS) data developed by parks, other agencies or organizations; national or international taxonomic or other classification systems; climate, air quality, or hydrologic data collected or assembled by regional or national entities.

4. *Data acquired and maintained by network parks that NCPN assists in managing.*
Because of the lack of data management expertise in many network parks, NCPN provides data management assistance for high-priority data sets or those that may benefit from standardized procedures. Examples include: a multi-park database for rare plant data; data sets of legacy natural resource monitoring data; and data on exotic invasive plant species.

These above categories can contain one or more of the following data formats:

- hard-copy documents (e.g., reports, field notes, survey forms, maps, references, administrative documents)
- physical objects (e.g., specimens, samples, photographs, slides)
- electronic text files (e.g., Word files, email, websites)
- electronic tabular data (e.g., databases, spreadsheets, tables, delimited files)
- spatial data (e.g., shapefiles, coverages, remote-sensing data)
- miscellaneous electronic files (images, sounds, other files with proprietary formats)

Each of these data formats has specific requirements for ongoing management and maintenance, which are addressed in the Data Management Plan.

Data Management and the Project Lifecycle

Inventory and monitoring projects are typically divided into five broad stages: planning and approval; design and testing; implementation; product integration; and evaluation and closure (Figure 1). During all stages data management staff collaborate closely with project leaders and participants.

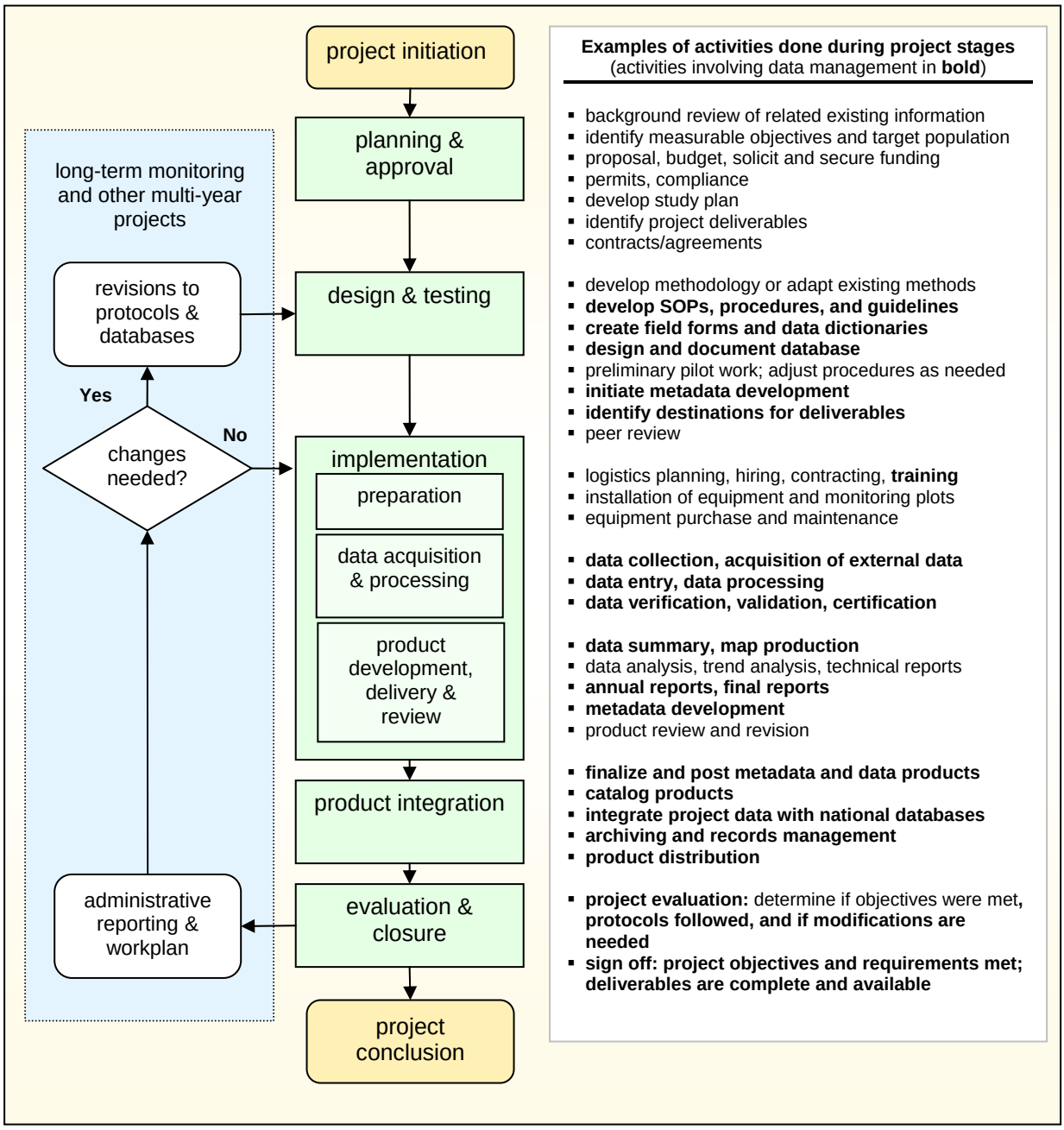

Figure 1. Model of data lifecycle stages and associated activities for the Northern Colorado Plateau Network.

Specific data management procedures corresponding to these stages are described in the chapters of the Data Management Plan. Building upon the data management framework presented in chapters 1 through 4, Chapter 5 is devoted to data acquisition and processing, and Chapter 6 provides a framework for verifying and validating data that

have been collected and entered into databases. Dataset documentation is the subject of Chapter 7, data reporting and analysis is presented in Chapter 8, and data dissemination, including issues such as data ownership and compliance with the Freedom of Information Act (FOIA), are addressed in Chapter 9. Chapter 10 provides a framework for the long-term maintenance, storage, and security of NCPN data.

Water Quality Data

The water quality component of the Natural Resource Challenge requires that networks archive all water quality data collected as part of the monitoring program in a STORET (STORage and RETrieval) database maintained by the NPS Water Resources Division (WRD). NCPN has developed an MS-Access database (NCPN H_2O) that consolidates available water quality data collected in and near the 16 NCPN park units. Associated with this database are water quality standards assessment tools that allow comparisons of historical and current data with applicable state standards. NCPN will maintain this database and integrate new data collected so it can serve as an ongoing tool for the network's long-term water quality monitoring and analysis needs.

On an annual basis NCPN will compile and format new water quality data from NCPN H_2O into an electronic data deliverable (EDD) that is compatible with WRD-STORET. WRD will ensure that content is transferred to the Environmental Protection Agency's STORET database (Figure 2).

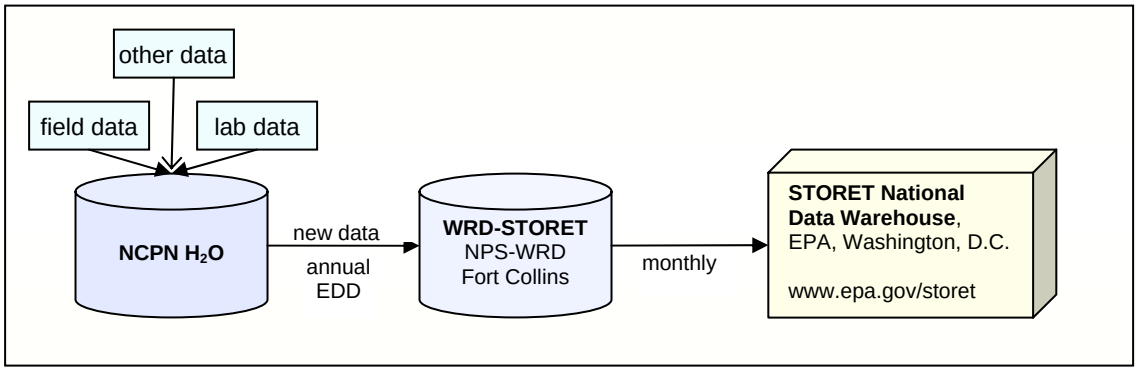

Figure 2. Simplified data flow diagram for water quality data.

Data Management Plan Maintenance

The NCPN approach is to maintain a Data Management Plan that is useful to a broad audience, and that can provide guidance on data management practices at a number of different levels. NCPN will keep the plan simple, flexible, and evolving, and include data users in the decision-making process whenever possible.

The document has undergone an initial prescribed review process that included both an internal network review (i.e., by members of the technical committee and network staff), and a service-wide review that involved the regional data/GIS coordinator, data management staff from the WASO I&M Program, and other network data managers.

NCPN will update the plan to ensure that it reflects accurately the network's current standards and practices. Recommendations for changes can be forwarded to the network data manager by any interested party or user of network inventory and monitoring data

(e.g., park resource managers, project leaders, technicians, superintendents, external users). These recommendations will be discussed by data management and network staff and actions decided upon. Simple changes can be made immediately in the document, while substantive changes will be made during version updates.

The most current version of the plan is available on the NCPN website (www.nature.nps.gov/im/units/ncpn).